

Psychological Monographs

General and Applied

No. 379
1954

Liberman, Delattre, Cooper, Gerstman

Vol. 68
No. 8



The Role of Consonant-Vowel Transitions
in the Perception
of the Stop and Nasal Consonants

By

Alvin M. Liberman, Pierre C. Delattre,
Franklin S. Cooper, and Louis J. Gerstman

Hashkins Laboratories, New York

Price, \$1.00

Edited by Herbert S. Conrad
Published by The American Psychological Association, Inc.

Psychological Monographs:

General and Applied

*Combining the Applied Psychology Monographs and the Archives of Psychology
with the Psychological Monograph*

Editor

HERBERT S. CONRAD

Department of Health, Education, and Welfare
Office of Education
Washington 25, D.C.

Managing Editor

LORRAINE BOUTHELET

Consulting Editors

DONALD E. BAKER

FRANK A. BEACHE

ROBERT G. BERNREUTER

WILLIAM A. BROWNELL

HAROLD E. BURTT

JERRY W. CARTER, JR.

CLYDE H. COOMBS

JOHN G. DARLEY

JOHN F. DASHIELL

EUGENIA HANFMANN

EDNA HEIDBREDER

HAROLD E. JONES

DONALD W. MACKINNON

LORRIN A. RIGGS

CARL R. ROGERS

SAUL ROSENZWEIG

ROSS STAGNER

PERCIVAL M. SYMONDS

JOSEPH TIFFIN

LEDYARD R. TUCKER

JOSEPH ZUBIN

MANUSCRIPTS should be sent to the Editor.

Because of lack of space, the *Psychological Monographs* can print only the original or advanced contribution of the author. Background and bibliographic materials must, in general, be totally excluded or kept to an irreducible minimum. Statistical tables should be used to present only the most important of the statistical data or evidence.

The first page of the manuscript should contain the title of the paper, the author's name, and his institutional connection (or his city of residence). Acknowledgments should be kept brief, and appear as a *footnote* on the first page. No table of contents need be included. For other directions or suggestions on the preparation of manuscripts, see: CONRAD, H. & Preparation of manuscripts for publication as monographs. *J. Psychol.*, 1948, 26, 447-459.

Correspondence concerning business matters (such as author's fees, subscriptions and sales, change of address, etc.) should be addressed to the American Psychological Association, Inc., 1209 Sixteenth St., N.W., Washington 6, D.C. Address changes must arrive by the 5th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee third-class forwarding postage.

Copyright, 1954, by THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Psychological Monographs: General and Applied

The Role of Consonant-Vowel Transitions in the Perception
of the Stop and Nasal Consonants¹Alvin M. Liberman,² Pierre C. Delattre,³ Franklin S. Cooper,
and Louis J. Gerstman*Haskins Laboratories, New York*

IN SPECTROGRAMS of stop consonant-plus-vowel syllables one commonly sees several acoustic variables that might conceivably be important in the auditory identification of the stop consonant phones. One such variable is a short burst of noise, found near the beginning of the syllable, as in Fig. 1, and presumed to be the acoustic counterpart of the articulatory explosion. By preparing hand-drawn spectrographic patterns of burst-plus-vowel and then converting these patterns into sound, we found in an earlier experiment (6) that the frequency position of the burst could serve as a cue, though not necessarily as a completely adequate one, for distinguishing among *p*, *t*, and *k*.

Bursts above 3000 cps were, in general, judged to be *t*. Below that level, the perception of the burst was determined by its frequency position in relation to the vowel with which it was paired: the burst was heard as *k* when it lay at or slightly above the second formant of the following vowel; otherwise, it was identified as *p*. The effect of the vowel on the perception of the burst was shown most strikingly in the case of one burst, centered at 1440 cps, which was heard as *p* before *i* and *u*, but as *k* before *a*.

¹ This research was supported in part by the Carnegie Corporation of New York and in part by the Department of Defense in connection with Contract DA49-170-SC-773. The first of the two experiments reported here was described at the 1952 meeting of the American Psychological Association; also, it was summarized in a discussion of related research at the 1952 Conference on Speech Analysis (2).

² Also at the University of Connecticut.

³ Also at the University of Colorado.

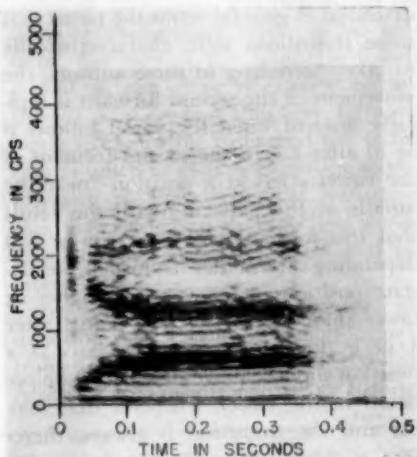


FIG. 1. Spectrogram of the Syllable *ga*, Showing a Burst and a Transition.

A second possible cue to the perception of the stops lies in the transition between consonant and vowel, seen in Fig. 1 as a curvature of the formants⁴ during the vowel onset. Such shifts in the frequencies of the vowel formants presumably reflect the articulatory move-

⁴ Formant: A relatively high concentration of acoustic energy within a restricted frequency region. In spectrograms the formants appear as dark bands whose general orientation is parallel to the horizontal (time) axis of the graph. Typically, three or more formants are seen, as, for example, in Fig. 1; these several formants are conventionally referred to by number, formant one being the lowest in frequency and in position on the spectrogram, formant two the next higher, and so on.

ments that are made in going from one position to another; one expects, then, to find these shifts in the region where two phones join.⁵

In their discussion of the cues that might be used in "reading" speech spectrograms, Potter, Kopp, and Green (7, pp. 81-103) have noted the transitions between stop consonant and vowel, especially in the second formant, and have described in general terms the forms that these transitions seem characteristically to take. According to these authors, the movement of the second formant is typically upward when the vowel follows *p* or *b*; after *t* or *d* the second formant of the vowel starts at a position "near the middle of the pattern," with the result that this formant will then rise or fall depending on whether its normal, steady-state position in the vowel is higher or lower than the *t-d* starting point; after *k* or *g* the second formant starts at a position slightly above its steady-state position in the vowel, wherever that may be, and the transition is always, therefore, a relatively small shift downward. Joos (5, pp. 121-125), also, has noted that the transitions are characteristically different for the various syllabic combinations of stop and vowel, and, in addition, has made explicit the assumption that the transitions may well be important cues for the perception of the stops. Without this latter assumption, he points out, one may have difficulty in explaining how listeners distinguish among the stop sounds, since the explo-

sive portions are sometimes of very low intensity.

Further analysis of spectrograms may result in a more nearly precise description of the typical patterns of transition for the various combinations of stop consonant and vowel. It will be no less necessary, however, to isolate the transitions experimentally if we are to determine whether they are perceptual cues or nulls, since the transitional movements do not occur independently of other possible cues in spectrograms of actual speech. To find whether or not the transitions can, in fact, enable a listener to distinguish among the stop consonants, we have, in the first experiment to be reported here, varied the direction and extent of second-formant transitions in highly simplified synthetic syllables and presented the resulting sounds to a group of listeners for identification as *b*, *d*, or *g* and, separately, as *p*, *t*, or *k*.

Since the nasal consonants *m*, *n*, and *ŋ* are closely related in articulatory terms to the voiced and unvoiced stops—*p-b-m* are all articulated by the lips, *t-d-n* by the tip of the tongue against the alveols, and *k-g-ŋ* by the hump of the tongue against the velum—we might guess that the transitional cues for *p-t-k* or *b-d-g* would also serve to distinguish among *m-n-ŋ*. A second experiment was carried out, then, to determine whether the variable second-formant transitions of the first experiment could function as cues for the perception of the nasal consonants. For that purpose we added to the patterns of vowel-plus-transition a certain neutral and constant resonance that had been found to impart to these patterns, as heard, the color or character of nasal consonants as a class. The sounds produced from these patterns were presented to subjects for judgment as *m*, *n*, or *ŋ*.

⁵We do not mean to imply that frequency shifts occur only between regions of steady-state resonance. It is, in fact, not unusual, especially in spectrograms of connected speech, to find that the formants are in almost constant movement. We shall deal here, however, with consonant-vowel syllables in which the vowel does assume a steady state following the transitional movement at the vowel onset.

EXPERIMENT I: STOP CONSONANTS

Apparatus. All the stimuli of this experiment were produced by using a special purpose playback to convert hand-painted spectrograms into sound. This playback, which has been described in earlier papers (1, 3), produces 50 bands of light modulated at harmonically related frequencies that range from a fundamental of 120 cps through the fiftieth harmonic at 6000 cps. The modulated light beams are arranged to match the frequency scale of the spectrogram. Thus, when a spectrogram, painted in white on a transparent base, is passed under the lights, the painted portions reflect to a phototube those beams whose modulation frequencies correspond to the position of the paint on the vertical (frequency) axis of the spectrogram.

Stimuli. As shown in Fig. 2, the second-formant transitions, which constituted the experimental variable of this study, differed in direction and extent.

From the frequency at which the second formant begins to the frequency at which it levels off, the transitions vary in steps of one harmonic (120 cps) from a point four harmonics (480 cps) below the center of the steady-state portion of the second formant to a point six harmonics (720 cps) above the second formant. This range of transitions was judged, on the basis of exploratory work, to be sufficient. For two of the synthetic vowels, *o* and *u*, the close proximity of first (lower) and second (higher) formants made it impossible to extend the transition as much as 480 cps below the second formant; in these cases the transitions that rise from points below the second formant were varied in four half-harmonic steps.

For convenience in reference, the direction and extent of second-formant transitions will be indicated as in section A of Fig. 2. Transitions that go through

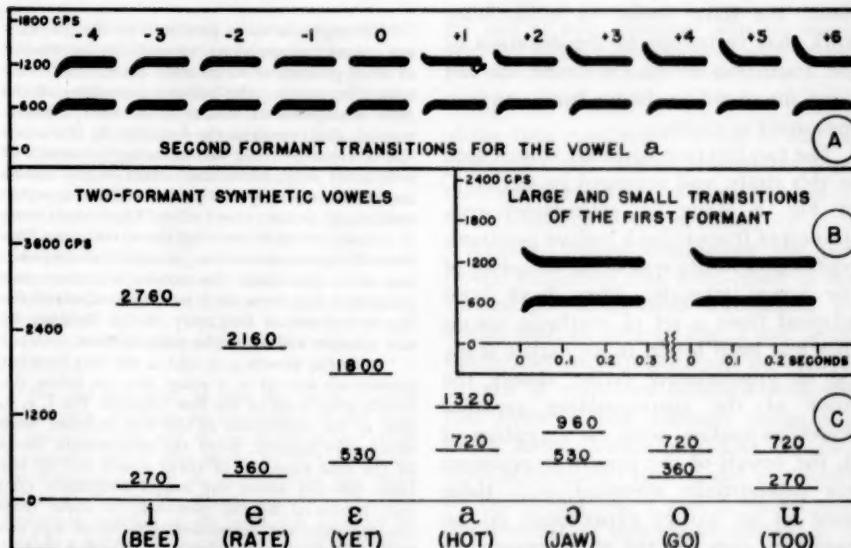


FIG. 2. The Stimuli of Experiment I.

a frequency range lying below the steady-state formant of the synthetic vowel will be called "minus" (-); those that cover a range of frequencies above the steady-state formant will be called "plus" (+). The extent of a transition will be given by the number of harmonics (of the 120-cps fundamental) through which the formant moves before arriving at its steady-state position. A transition of -3, for example, is one that goes through the first three harmonics below the steady state of the vowel.

In determining the curvature of the transitions, we tried simply to approximate the transitions we have seen in spectrograms of actual speech. The duration of the transition, i.e., the time interval between the beginning and end of the frequency shift, varied linearly with the size of the transition, from 0.04 sec. for a transition of +1 or -1, to 0.08 sec. for a transition of +6. This made it possible to keep the shape of the transition (as judged by eye) roughly constant. We have found in exploratory work that variations in the duration of the transition or in curvature do not cause the sound to change from one stop consonant to another.

The two-formant synthetic vowels used in this study, and arranged in section C of Fig. 2 along an articulatory continuum of front-to-back tongue positions, represent a rather systematic sampling of the vowel triangle. They have been adapted from a set of synthetic vowels that had been found in an earlier study (4) to approximate rather closely the color of the corresponding cardinal vowels as spoken; with the exception of *u*, the vowels of the present experiment are substantially identical with those used in an earlier experiment (6) on bursts as cues for the stop consonants. The frequency extent of each formant is

three harmonics of the 120-cps fundamental. The value that is to be found just above each formant in section C of Fig. 2 gives the frequency that corresponds approximately to the center of the formant.⁶ Each pattern of transition-plus-vowel has a total duration of 0.3 sec.

In our essentially exploratory attempts to produce acceptable stop consonants with nothing more than a transition and a schematic vowel, we had adopted for the first formant the constant minus transition seen in section A of Fig. 2.⁷ This first-formant transition seemed to increase the realism and identifiability of the sounds, but it imparted to all of them a rather strong voiced quality; that is, it made them sound much more like *b-d-g* than *p-t-k*. Although we were not concerned in this study to isolate the cues that distinguish the voiced stops from their voiceless counterparts, we did wish to investigate the role of second-formant transitions in both classes of sounds, and, on the assumption that phonetically

⁶ Although the tones produced by the playback are spaced 120 cps apart, an auditory impression of finer gradations of formant pitch can be obtained by varying the relative intensities of the three contiguous harmonics (of the 120-cps fundamental) that comprise the formant. In this study the intensities of the two outlying harmonics of a formant were sometimes intentionally unbalanced in an attempt to produce closer approximations to correct vowel color. The unbalancing is accomplished by varying the extent to which the white paint covers the "channel" corresponding to a particular harmonic. Wherever this procedure has been used, we have estimated the equivalent center frequency of the formant by the relative widths of the painted lines.

⁷ With the vowels *e*, *a*, and *ɔ*, the first-formant transitions started at a point 360 cps below the steady-state level of the first formant. For *i*, *e*, *o*, and *u*, the transitions of the first formant were necessarily smaller, since the steady-state levels of the first formants of these vowels are all less than 360 cps above the lowest frequency (120 cps) produced by the playback; in these cases the first-formant transitions started at 240 cps below the steady-state level for *e* and *o*, and at 120 cps below the steady state for *i* and *u*.

naive listeners might have difficulty in trying to identify the voiced sounds as voiceless, we thought it wise to try to "unvoice" the sounds before presenting them for judgment as *p-t-k*. The closest approximation we could achieve, without adding the burst of noise that appears to characterize *p-t-k*, was obtained by considerably reducing the transition of the first formant.* This reduces somewhat the impression of voicing; it makes the stops resemble the unaspirated voiceless stops used, for example, by a native speaker of French, but it does not succeed in producing the clearly unvoiced quality typical of aspirated *p-t-k* in an American pronunciation. The "unvoiced" and "voiced" types of first-formant transition are illustrated in section B of Fig. 2.

There were, then, two sets of stimuli that were identical in all respects, except that in one set the minus transition of the first formant was relatively large and in the other very small.

Presentation of stimuli. A total of 77 stimuli (11 transitions times 7 schematic vowels) was used for each of the two sets of test patterns (voiced and unvoiced). These 77 stimuli were recorded from the playback onto magnetic tape, and then spliced into a random order, subject to the restrictions that in each successive group of 11 sounds each transition appear once and only once, and that each vowel appear at least once in each group but never more than twice and never in immediate succession.

In the final test tape the sound stimuli were arranged in such a way that each stimulus would be presented and then repeated after an interval of 0.9 sec., with an interval of 6 sec. between successive pairs of identical stimuli. The

* We did not wish to add the burst because it not only produces an impression of the class of unvoiced stops, but, as was shown in an earlier experiment (6), it also serves, by its position on the frequency scale, to differentiate the stops within that class. We should note here that the frequency position of the burst can probably be used as a cue for distinguishing among the voiced stops also. This will, of course, require the addition of certain "voicing" constants.

latter interval provided sufficient time for *S* to make and record his judgment of one stimulus before being presented with the next. A rest period of 15 sec. was interpolated between successive groups of 11 stimuli.

The "voiced" patterns (i.e., those with large transitions of the first formant) were presented to one group of 33 *Ss* for judgment as *b*, *d*, or *g*; the "unvoiced" patterns (i.e., those with small transitions of the first formant) were presented to a second group of 33 *Ss* for judgment as *p*, *t*, or *k*. Then, to exhaust all combinations of the two kinds of first-formant transition and the two kinds of judgment, we recruited two additional groups of 33 *Ss* each and asked one of these groups to judge the patterns with the large first-formant transitions as *p*, *t*, or *k*, and the other to judge the patterns with the small transitions of the first formant as *b*, *d*, or *g*.

In all cases *S* was asked to make an identification of each stimulus, even when he felt that his judgment was only a guess. The range of identifications permitted was limited to *p*, *t*, or *k* for two groups of *Ss*, and to *b*, *d*, or *g* for the other two groups.

The entire stimulus series was presented twice for each *S*. Thus, each *S* made a total of 154 judgments.

Before *Ss* began to record their judgments, they were asked to listen to the first group of 11 stimuli in order that they might become somewhat acquainted with the nature of the sounds and the format of the experiment. For all *Ss* the opportunity to hear these 11 stimuli constituted the sum of their experience with the sounds produced by the pattern playback prior to their participation in this experiment.

Subjects. A total of 132 *Ss* (33 in each of the four conditions) served in the experiment. All were volunteers from undergraduate and graduate courses at the University of Connecticut.

Results. Before considering the particular responses that were made to the various second-formant transitions, we ought, first, to note in Fig. 3 and 4 the similarities and differences in the general pattern of responses obtained with the two types of first formant (large and small minus transitions) and the two types of judgment (voiced and unvoiced). By comparing Fig. 3 and 4, we see that the amount of agreement among *Ss* is somewhat greater when the first formant has the larger minus transition. It will

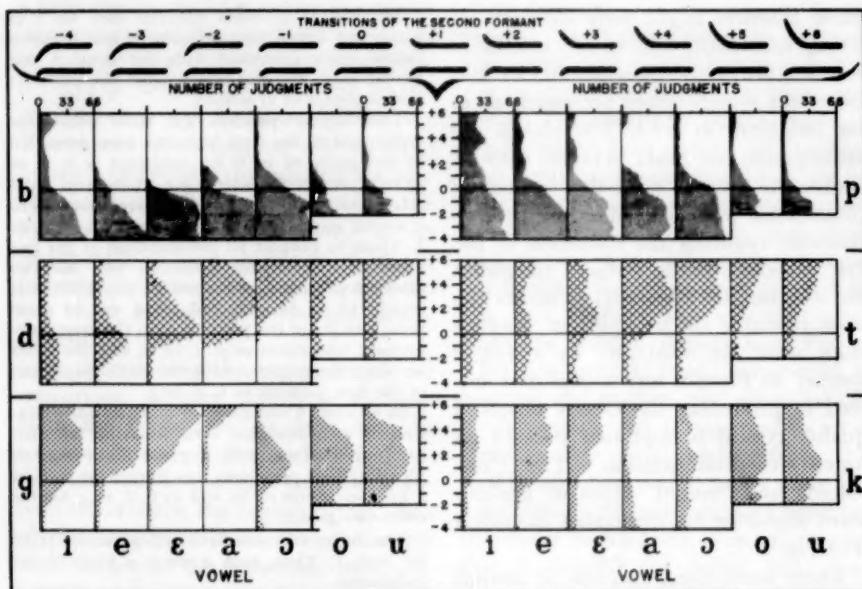


FIG. 3. Responses to the Variable Transitions of the Second Formant Obtained with Large Transitions of the First Formant. Within each small rectangle, the number of times a stimulus was judged as *b*, *d*, or *g* (or as *p*, *t*, or *k*) is plotted on the horizontal axis as a function of the direction and extent of the second-formant transition (shown on the vertical axis and illustrated at the top of the figure). The data were obtained from two groups of 33 Ss each. One group judged the stimuli as *b*, *d*, or *g*, the other as *p*, *t*, or *k*. Each S made two judgments of each stimulus. The vowels with which the second-formant transitions were paired are arranged along an articulatory continuum of tongue position from front to back. Because of the close proximity of the first and second formants of *o* and *u*, the minus transitions for these vowels were not extended beyond a value of two.

be remembered in this connection that we had adopted the relatively straight first formant in an attempt to reduce the impression of voicing; that is, to make the synthetic stops sound less like *b-d-g* and somewhat more like *p-t-k*. Clearly, we did not succeed by this means in increasing the amount of agreement among Ss who tried to identify these stimuli as *p-t-k*.

We have noted earlier that it was not the purpose of this study to find the cues that distinguish voiced from unvoiced stops, and it should be emphasized here that we have not yet investigated this matter, except cursorily and in a few isolated cases. On the basis of such evidence as is now available, however, we believe that the voiced-voiceless distinction is more easily made

if one is free, as we were not in this particular experiment, to use both burst and transition. The proper combination of burst and transition tends to produce an impression of the general class of unvoiced stops, the particular identity of the stop (i.e., *p*, *t*, or *k*) being determined both by the frequency position of the burst and the nature of the transition. It will presumably be possible, then, to obtain the voiced counterparts by making certain constant changes in the pattern, as, for example, by varying the time interval between burst and transition, or by adding constant "markers," such as a "voice bar" (a tone of 120 or 240 cps sounding simultaneously with the burst). Thus, a particular combination of burst and second-formant transition would, with any given vowel, be used for the synthesis of either *p* or *b*, a second such combination would be used for *t* or *d*, and a third for *k* or *g*; one could expect to shift back and forth between the voiced and voiceless stops without making significant changes in the frequency position of the

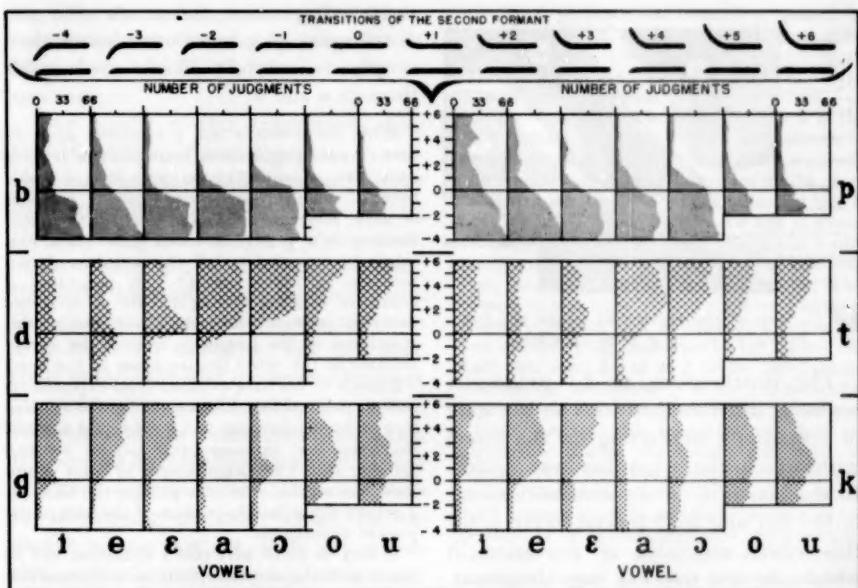


FIG. 4. Responses to the Variable Transitions of the Second Formant Obtained with Small Transitions of the First Formant. The data are displayed as in Fig. 2.

burst or in the direction and extent of the second-formant transition.

With large or with small transitions of the first formant (Fig. 3 or 4) we see that the amount of agreement is, in general, slightly greater when Ss are trying to identify the sounds as *b-d-g*. (This is not surprising, perhaps, in view of the fact that all of the stimuli sounded quite voiced.) It is nonetheless clear that the various second-formant transitions produce the same general pattern of responses, whether the stimuli that contain them are judged as *b-d-g* or as *p-t-k*. For the purposes of this paper, then, we shall consider that the second-formant transitions have essentially the same effect within each of the two classes (voiced and unvoiced) of stop consonants, and we shall treat the results as if each of the pairs, *p-b*, *t-d*, and *k-g*,

were a single sound.

The response distributions of Fig. 3 and 4 show, in general, that *b* (or *p*) was heard when the second formants of the vowels had minus transitions (i.e., transitions that extend into a frequency region lower than the frequency of the steady-state portion of the formant). When these minus transitions were presented with the vowels *i*, *e*, *ɛ*, *a*, and *ɔ*, there was considerable, and in some cases complete, agreement among Ss in identifying the sounds as *b-p*. With the vowels *o* and *u*, on the other hand, these same transitions elicited few responses of *b-p* (in relation to *g-k* and *d-t*), but there is, nevertheless, a significant similarity in the pattern of judgments as between *o-u* and *i-e-ɛ-a-ɔ* in that the bulk of *b-p* judgments occurs, for all these vowels, within the range of minus transitions.

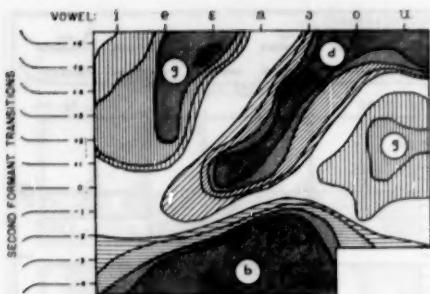


FIG. 5. Map of the regions in which the judgments *b*, *d*, or *g* were dominant.

The distributions of *d-t* judgments center at different positions on the scale of transitions, depending on the vowel with which the transition was paired. In the case of *i* the *d-t* distributions are rather flat and low; indeed, there is for this vowel no value of transition at which the *d-t* response was dominant. For the remaining vowels, however, there is a rather strong preference for identifying *d* (or *t*) within some relatively narrow range of transitions, starting in the vicinity of a zero transition for the vowels *e* and *ɛ*, and progressing to larger plus transitions through the vowel series *a-ɔ-o-u*.

The distributions of *g-k* judgments would appear, for all plus transitions at least, to be the inverse of the *d-t* distributions; that is, the *g-k* judgments occur wherever the *d-t* judgments do not. Considered in their own right, the distributions of *g-k* responses show that the extreme plus transitions were heard as *g-k* in the vowel series *i* to *a*. The smaller plus transitions were also heard as *g-k* with *i* and *e*, but at *e* only the transitions of +4, +5, and +6 were judged very often as *g* or *k*, and at *a* such *g-k* responses as did occur were made, for the most part, to the extreme transition of +6. With *ɔ*, the *g-k* responses center at

the small positive transitions and the distribution of $g\text{-}k$ responses seems then simply to grow in height and width through o and u .

With the vowels *e* and *ɛ* relatively good *d*'s and *t*'s were produced by transitions of zero, or near zero, extent. This suggests that *d* and *t* have a characteristic second-formant position, or locus, somewhere near the level of the second formant of *e* or *ɛ*, from which point transitions might be expected to fall or rise to the second formants of other vowels.⁹ This possibility is consistent with our finding that the *d-t* responses occurred primarily to progressively larger plus transitions as the frequency level of the second formant of the vowel became lower in the series *a* through *u*. We have evidence from experiments now in progress that adds support to the assumption of fixed consonant loci for *d-t*, and suggests that there may be comparable loci for *b-p* and for *g-k*, also. The establishment of these consonant loci would, of course, provide the basis for a greatly simplified description of the data of the present experiment.

It may be noted here that a consonant will be heard with the second formant at zero transition only when the first formant has some degree of minus transition. When both formants are straight, one hears nothing but the vowel.

Figure 5, which is intended to be a broader and less detailed representation of the results, is derived from the data in the left half of Fig. 3 (stimulus patterns with the large transitions of the first formant, judged as *b-d-g*). For the purposes of Fig. 5, a particular response was taken as "dominant" when (for any stimulus) the difference in the number of judgments between the most numerous and the next most numerous response exceeded 20 per cent of the sum of all responses. This constitutes the lowest degree of dominance, and is indicated in the figure by the lightest shading.

⁹ Potter, Kopp, and Green (7, pp. 81-103) have inferred from spectrograms the existence of a "hub," or second formant, for each of the stop consonants. They assume that the frequency position of this hub is fixed for *p* and *t*, being always at a relatively low frequency for *p* and a relatively high frequency for *t*. The hub of *k* is assumed to vary according to the following vowel.

ing. The three darker shadings of Fig. 5 correspond to increasing degrees of dominance (40, 60, and 80 per cent, respectively).

The results of this experiment show that the direction and degree of second-formant transitions can serve as cues for the aurally perceived distinctions among the stop consonants.¹⁰ In judging some of the stimuli all, or almost all, Ss made the same identifications. In other cases, the amount of agreement was considerably less than complete, but still sufficient to indicate that the second-formant transition has considerable cue value. However, certain cases remain in which the second-formant transition does not appear to provide an adequate basis for identifying the stop: with *i*, for example, we do not get, with any transition, a clearly dominant *t* or *d* response.

There is evidence from preliminary investigations that an increase in the identifiability of the synthetic stops will result from the inclusion of appropriate transitions of the *third* formant. It is quite clear, however, that the third-formant transitions are, in general, considerably less important for the perception of the stop consonant than are transitions of the second formant.

EXPERIMENT II: NASAL CONSONANTS

Procedure. The stimuli of Experiment II were identical with those of Experiment I, except that (a) the transitions were placed at the ends of the syllables

¹⁰ It appears that the *duration* of transition is also a cue for distinguishing among speech sounds, not within the class of stop consonants, but between this class and others. Thus, increasing the transition time beyond the largest value (0.08 sec.) used in the present experiment causes some of the stop consonants to be transformed first into semivowels and then, with further increases in duration of transition, into vowels of changing color. In the case of a +6 transition

We have prepared and listened to a series of patterns in which, for each of the seven schematic vowels (*i*, *e*, *ɛ*, *a*, *ɔ*, *o*, and *u*), third-formant transitions of -3, 0, and +3 were added to each of nine second-formant transitions (from -4 through 0, to +4). When -3 transitions of the third formant are added to minus transitions of the second formant, *b* is heard with all vowels. The addition of the -3 transition in the third formant seems in these cases to improve the *b*, particularly with the back vowels *o* and *u*, where, according to the data of Fig. 3 and 4, the *b* produced without third-formant transition is relatively poor. Adding a -3 transition of the third formant to plus transitions of the second formant produces, in general, a *g* impression, and in the cases of those plus transitions of the second formant that were heard as *g* in the two-formant patterns, the identifiability of the *g* is somewhat improved. As might be expected, the addition of a third formant with zero transition does not appreciably change the sound; one hears essentially what was heard with the two-formant version. The addition of a +3 transition of the third formant tends in general to produce an impression of *d*. Combining this third-formant transition with the appropriate second-formant transition creates a *d* that seems significantly better than the best that could be produced with first- and second-formant transitions alone.

Reference to the results of the earlier study on bursts will show that bursts are often effective just where the second-formant transitions alone fail. We should expect, then, that adding an appropriate burst to the transitions will further reduce the number of cases in which the response is equivocal.

rather than at the beginnings, (b) neutral, constant resonances were added after the transitions, and (c) the first formants had no transitions, instead of

with *e*, for example, the sound will be heard as *ge*, then *je*, and, finally, as *ie*, as the duration of the transition is progressively lengthened. If the transition time is reduced below the lowest value used in the present experiment (0.04 sec.), a point is reached, eventually, at which the perception begins to change in various and complex ways; these changes are probably attributable to the fact that at very short durations the transitions are so abrupt as to be, in effect, bursts.

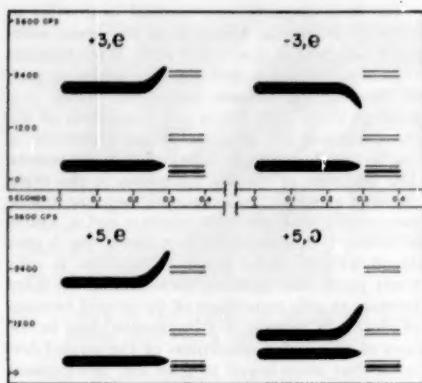


FIG. 6. Examples of the Stimuli Used in Experiment II.

the constant minus transitions of Experiment I. It was considered advisable to put the nasal consonants (*m*, *n*, *ŋ*) at the ends of the syllables because *ŋ* does not occur in the initial position in English. The purpose of the neutral resonance was to add to all the sounds the nasal quality characteristic of *m*, *n*, *ŋ* as a class. A straight first formant was used throughout because it had seemed in exploratory work that the best nasal consonants were produced in this way.

It will be seen in the patterns of Fig. 6 that the nasal resonance consists of three formants, centered at 240 cps, 1020 cps, and 2460 cps, and, also, that the formants of the nasal resonance are somewhat less intense than those used to produce the synthetic vowels. The particular frequency positions and intensities of these nasal "markers" were selected on the basis of exploratory experimentation, and were judged by the authors to produce rather indifferently the nasality of *m*, *n*, or *ŋ*.¹¹

The procedure for presenting the stimuli was identical with that of Experi-

¹¹ It is possible that the frequency positions of the "nasal" formants are, in actual speech,

ment I. The instructions to *Ss* were also exactly as they were in Experiment I, except, of course, that *Ss* were asked in Experiment II to identify the sounds as *m*, *n*, or *ŋ*.

Subjects. A total of 33 undergraduate and graduate students at the University of Connecticut judged the stimulus patterns. All *Ss* were without prior experience in judging or listening to the synthetic speech sounds produced by the playback.

Results. The response distributions of Fig. 7 show how the stimulus patterns of Experiment II, which contained the second-formant transitions of Experiment I plus a nasal resonance, were identified as *m*, *n*, or *ŋ*. To provide a ready comparison between the results of Experiments I and II we have reproduced in Fig. 7 the left half of Fig. 4, showing how the *b-d-g* judgments varied as a function of variations in the second-formant transitions (when the first formants had the small minus transitions that most closely approximate the straight first formants of Experiment II).

The minus transitions that were heard in the first experiment as *p* (or *b*) are heard here as *m*. The major difference between the stop and nasal consonants as cued by these minus transitions of the second formant would appear to be that the response to the stops is relatively strong (i.e., there is considerable agreement among *Ss* in identifying the minus transitions as *p-b*) for the vowels *i*, *e*, *ɛ*, *a*, *ɔ*, and relatively weak for *o* and *u*, while for the nasals the *m* response is

characteristically different for *m*, *n*, and *ŋ*; if so, the synthetic sounds will presumably be improved when these differences are included in the patterns. We have tried here simply to find a particular position of the formants that would produce a general nasal quality without strongly biasing the sound toward any one of the three nasal phones.

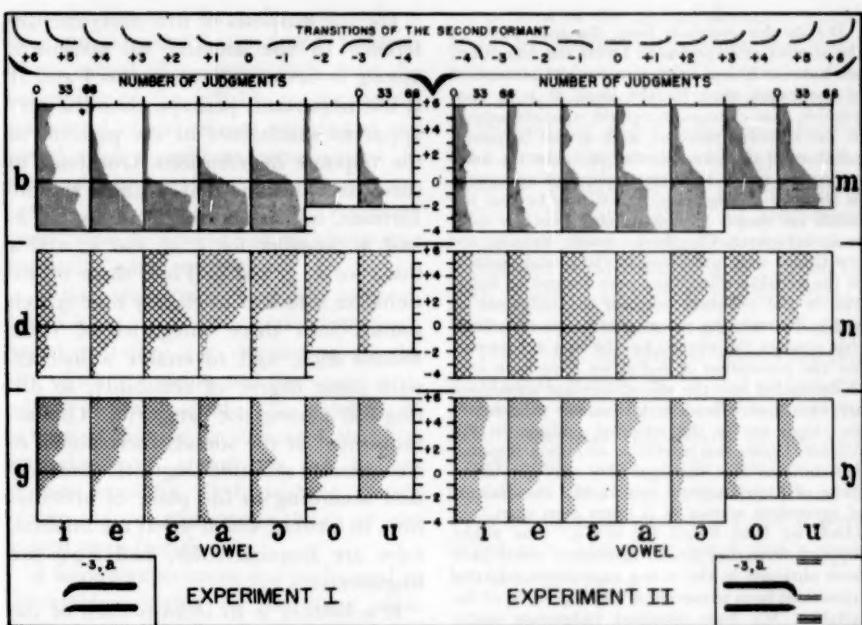


FIG. 7. Distributions of *m*, *n*, and η Judgments (Experiment II) and *b*, *d*, *g* Judgments (Experiment I). The data are displayed as in Fig. 2.

relatively weak with *i*, *e*, and strong with *ɛ*, *a*, *ɔ*, *o*, and *u*.

The distributions of *n* responses are quite flat for *i* and *e*; for the remaining vowels, the *n* responses clearly fall in the region of plus transitions and the bulk of the *n* responses tends to move toward higher values of plus transition from *ɛ* through *u*. As can be seen by comparison with the results of Experiment I, the distributions of *n* responses are very similar to those obtained for *d* (or *t*).

The distributions of η responses are relatively flat and low for all the vowels—it is only with *i* and *e* that any transition is judged more often as η than as *m* or *n*—but one sees nevertheless that the η responses correspond to the *k-g* responses of Experiment I in that they occur primarily in the region of plus

transitions. One can also see, perhaps, that there is a tendency, roughly comparable to that seen in the *k-g* distributions, for the bulk of the η responses to be displaced progressively toward the higher plus transitions for the vowels from *i* through *a*; beyond *a* the number of η responses is too few and the distributions are too nearly rectangular to permit detailed comparisons with *k-g*.

It is clear that the variable second-formant transitions can be cues for the perceived distinctions among *m*, *n*, and η . A comparison with the results of Experiment I, however, shows that the second-formant transitions were probably somewhat less effective as cues for the nasals than they were in providing a basis for distinguishing among the stops.

We do not conclude from the superiority of the stimuli of Experiment I that the transitions are necessarily less important for the perception of the nasals than for the stops. It is at least possible that changes in certain constant aspects of the stimulus patterns, such as the frequency positions of the formants that comprise the nasal resonance, might raise the amount of agreement in the *m-n-ŋ* judgments, and it may be that the nasals are simply less identifiable than the stops in actual speech. One must consider also that the transitions were presented in the initial position in the syllable when they were judged as *b-d-g*, but in the terminal position for judgment as *m-n-ŋ*. By reversing the magnetic tape recordings that contain the stimuli for the first experiment (on the perception of *b-d-g*) we have been able to determine how the second-formant transitions are identified when the transitions (and hence the stops) are in the terminal position in the syllable. Under this condition we obtain response distributions that have patterns very similar to those of Experiments I and II, but the amount of agreement among *Ss* is lower even than that which we have found for *m-n-ŋ*.¹² One might suppose, then, that greater agreement would have been obtained in the *m-n-ŋ* experiment had the transitions been presented at the beginning of the syllables. We have obtained judgments under this condition (by reversing the magnetic tape of Experiment II), but in this case we find no very large difference in the responses, either in regard to pattern or to amount of agreement. (The interpretation of this result is, of course, somewhat complicated by the fact that English-speaking *Ss* are not accustomed to hearing *ŋ* in the initial position.)

¹² It may be noted that the effect of reversing the magnetic tape recordings of the stimuli is to convert a rising transition in initial position to a falling transition in terminal position, although both are minus transitions in the terminology of this paper and both are identified (with more or less agreement) as the same consonant.

We have not yet explored the possibility that the terminal transitions must be somewhat different from initials, in intensity or rate of change, for example, if they are to be maximally effective.

For the purposes of this study the difference in the amount of agreement among *Ss* between Experiments I and II is less important, perhaps, than the very apparent similarities in the patterns of the response distributions. One finds in these distributions that a single second-formant transition can serve for *p*, *b*, and *m*, another for *t*, *d*, and *n*, and a third for *k*, *g*, and *ŋ*. Thus, three transitions are sufficient to classify nine speech sounds into three categories of three sounds each, and to enable a listener, with some degree of reliability, to distinguish among the categories. This arrangement of the sounds corresponds to a commonly accepted linguistic classification according to the place of articulation, in term of which *p-b-m* are bilabial, *t-d-n* are linguoalveolar, and *k-g-ŋ* are lingualvelar.

If a listener is to identify each of the nine sounds uniquely, he will need, in addition to the cues for place of articulation, some basis for determining whether a given sound belongs in the class of unvoiced stops (*p-t-k*), voiced stops (*b-d-g*), or nasal consonants (*m-n-ŋ*). It is possible that the distinctions among these three classes are effectively cued by a limited number of acoustic markers, such as voice bar and nasal resonance, each of which is constant within its class and characteristic of a manner, rather than a place, of articulation. The experiments reported here were not designed to test this possibility.

SUMMARY

Spectrograms of stop consonant-plus-vowel syllables characteristically show rapid transitional movements (frequency shifts) of the formants at the vowel onset. To determine whether the transitions (particularly those of the second for-

mant) are cues for the perceived distinctions among the stop consonants, two series of simplified, hand-painted spectrograms of transition-plus-vowel were prepared, then converted into sound by a special purpose playback and presented

to naive listeners for judgment as *b*, *d*, or *g* and, separately, as *p*, *t*, or *k*. In terms of the extent of frequency shift, from the beginning of the syllable to the steady-state level of the second formant of the vowel, the transitions were varied in steps of 120 cps from a point 480 cps below the second formant ("minus" transitions) to a point 720 cps above that formant ("plus" transitions).

Minus transitions were, in general, heard as *p* or *b*. Plus transitions were heard as *t-d* or *g-k*, depending on the size of the transition and the vowel with which it was paired. The amount of agreement among Ss indicated that the second-formant transitions can be important cues for distinguishing among either the voiceless stops (*p-t-k*) or the voiced stops (*b-d-g*).

A second experiment was performed to determine whether or not these same

transitions of the second formant would serve to distinguish among the nasal consonants (*m-n-ŋ*), which are related to the stops in that the buccal occlusion is bilabial for *m-p-b*, linguoalveolar for *n-t-d*, and linguovelar for *ŋ-k-g*. The stimulus patterns of the second experiment were identical in all important respects to those of the first, except that a constant nasal resonance was added to each pattern and that the transitions were placed at the ends of the syllables.

There was in general somewhat less agreement among Ss in the second experiment than there had been in the first. Otherwise the results of the two experiments were quite similar: the transitions that had served for *p* and *b* were heard as *m*, those for *t* and *d* were heard as *n*, and those for *k* and *g* were heard as *ŋ*.

REFERENCES

- COOPER, F. S. Spectrum analysis. *J. acoust. Soc. Amer.*, 1950, **22**, 761-762.
- COOPER, F. S., DELATTRE, P., LIBERMAN, A. M., BORST, J. M., & GERSTMAN, L. J. Some experiments on the perception of synthetic speech sounds. *J. acoust. Soc. Amer.*, 1952, **24**, 597-606.
- COOPER, F. S., LIBERMAN, A. M., & BORST, J. M. The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proc. Nat. Acad. Sci.*, 1951, **37**, 318-325.
- DELATTRE, P., LIBERMAN, A. M., COOPER, F. S., & GERSTMAN, L. J. An experimental study of the acoustic determinants of vowel color; observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word*, 1952, **8**, 195-210.
- JOOS, M. Acoustic phonetics. *Language, Suppl.*, 1948, **24**, 1-136.
- LIBERMAN, A. M., DELATTRE, P., & COOPER, F. S. The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *Amer. J. Psychol.*, 1952, **65**, 497-516.
- POTTER, R. K., KOPP, G. A., & GREEN, H. C. *Visible speech*. New York: Van Nostrand, 1947.

(Accepted for publication March 8, 1954)

GROUNDSANTA PUBLISHING COMPANY, NEW YORK